Investigation of the Exponential Distribution

Porter

March 2, 2016

Overview

In this investigational report there will be comparisons between a simulated exponential distribution and the theoretical exponential distribution. In order to have a meaningful comparison this report will run a thousand simulations and calculate the mean of 40 exponentials per simulation. Then the distribution of the simulations will be compared to the theoretical exponential distribution.

Simulations

The distribution that is explored in this report is the exponential distribution with a lambda of 0.2.

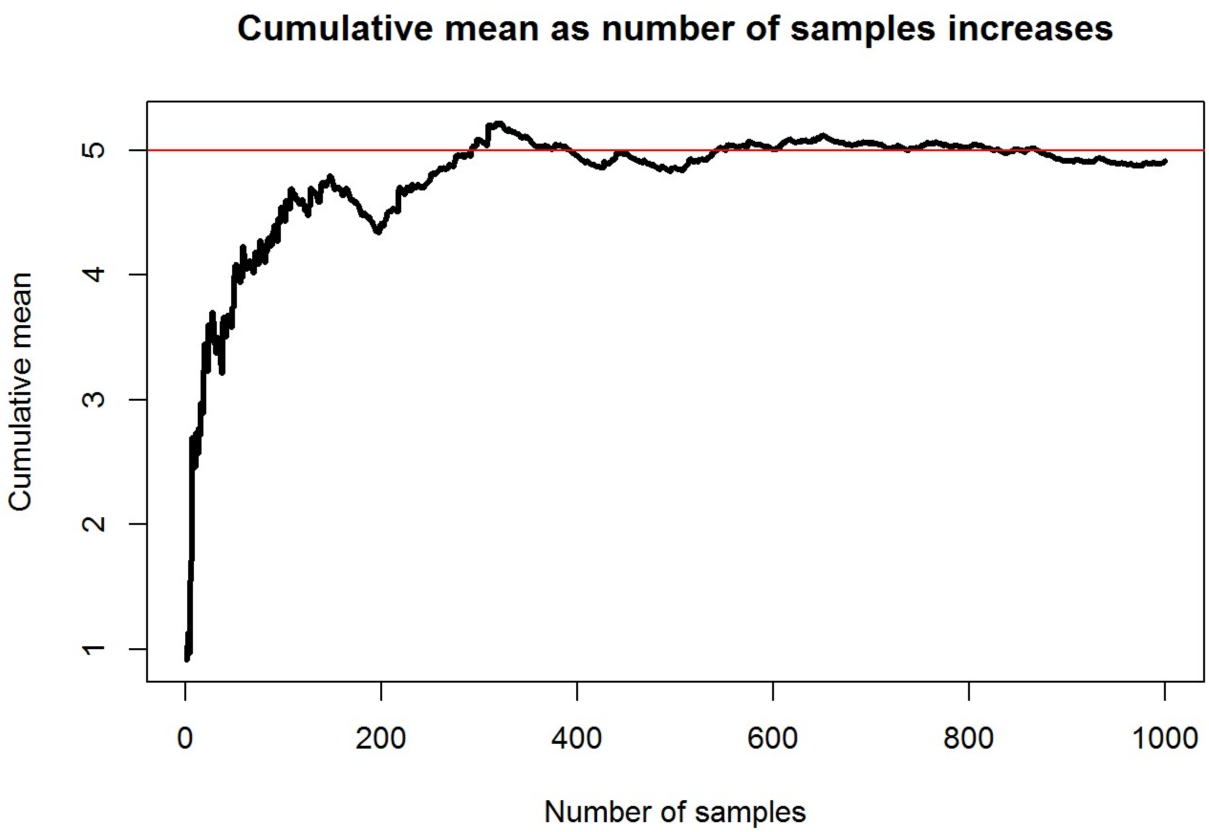
```r
n <- 1000
lambda <- 0.2
simulation_sample_size <- 40

theoretical_mean = 1/ lambda
theoretical_sd = 1/lambda
theoretical_variance = theoretical_sd^2

means = NULL
variances = NULL
cumsum_means = cumsum(rexp(n, lambda)) / (1:n)
for (i in 1:n) {
    exp_samples = rexp(simulation_sample_size, lambda)
    means = c(means, mean(exp_samples))
    variances = c(variances, sd(exp_samples)^2)
}

sample_mean = mean(means)
sample_variance = mean(variances)
```

The R code above runs the simulations. For 1,000 iterations it generates 40 samples from the exponential distribution and then calculates the mean and variance of those 40 samples. By the end of this code execution there will be a variable that stores 1,000 sample means and a variable that stores 1,000 sample variances from the exponential distribution. This code also generates the cumulative sum means so that we can see how the variability of the sample distribution changes as we obtain more samples. It also calculates the variance in the samples

Sample Mean vs Theoretical Mean

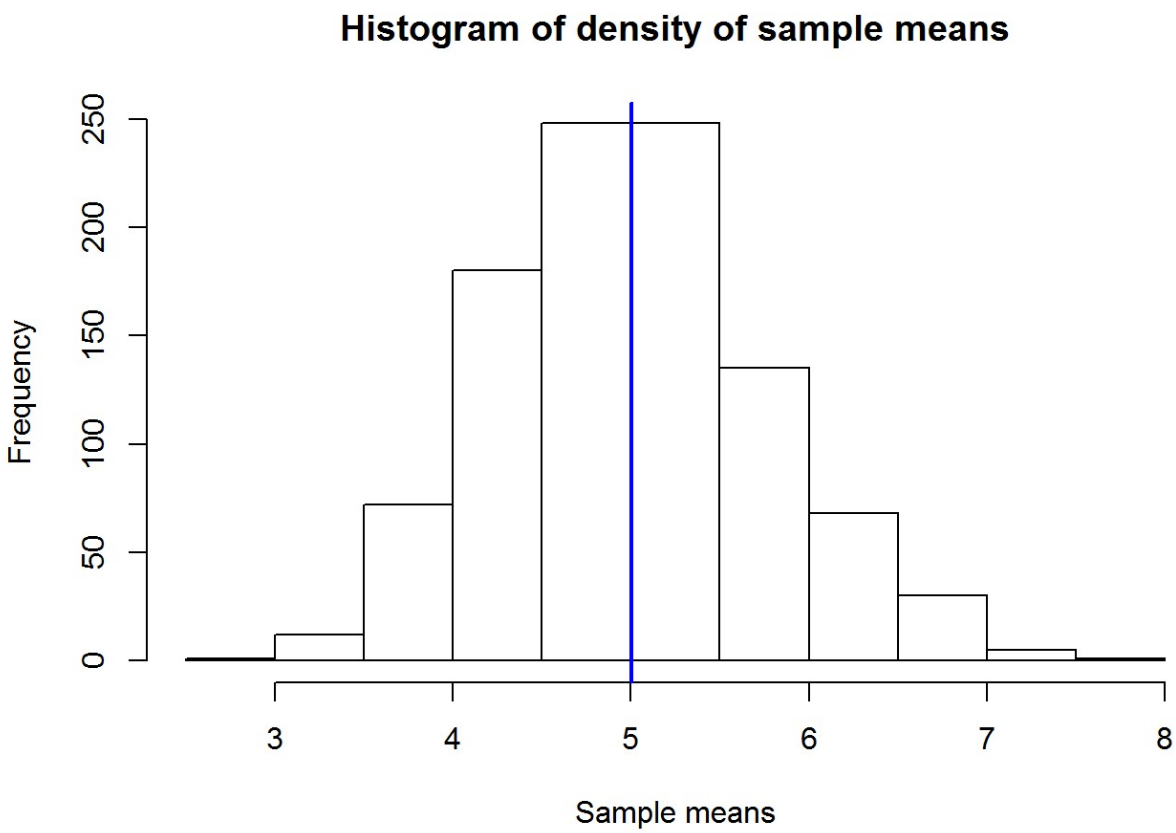
The code below will plot of the cumulative means as the sample size increases it starts to converge on approximately the theoretical mean. This adheres to the law of large numbers that says the average limits to what it's estimating. In this case the theoretical mean of the exponential distribution with a lambda of 0.2 is what is being estimated. The theoretical mean is shown as a horizontal red line in the plot. The theoretical mean is known to be $1/lambda$ where $lambda = 0.2$.

```
plot(c(1:n), cumsum_means, type = 'l', main = 'Cumulative mean as number of sam
ples increases',
     xlab = 'Number of samples', ylab = 'Cumulative mean', lwd=3)
abline(h = theoretical_mean, col='red')
```



The following code will show the histogram of the sample means and also draw the sample mean as a vertical line. As discussed previously the sample mean is $1/lambda$ which in this case is 5. The sample mean is we calculated from our simulations is 5.0047087. The following plot shows that the sample mean is quite close to the theoretical mean. In theory the more samples we took the better the sample mean would approximate the theoretical mean.

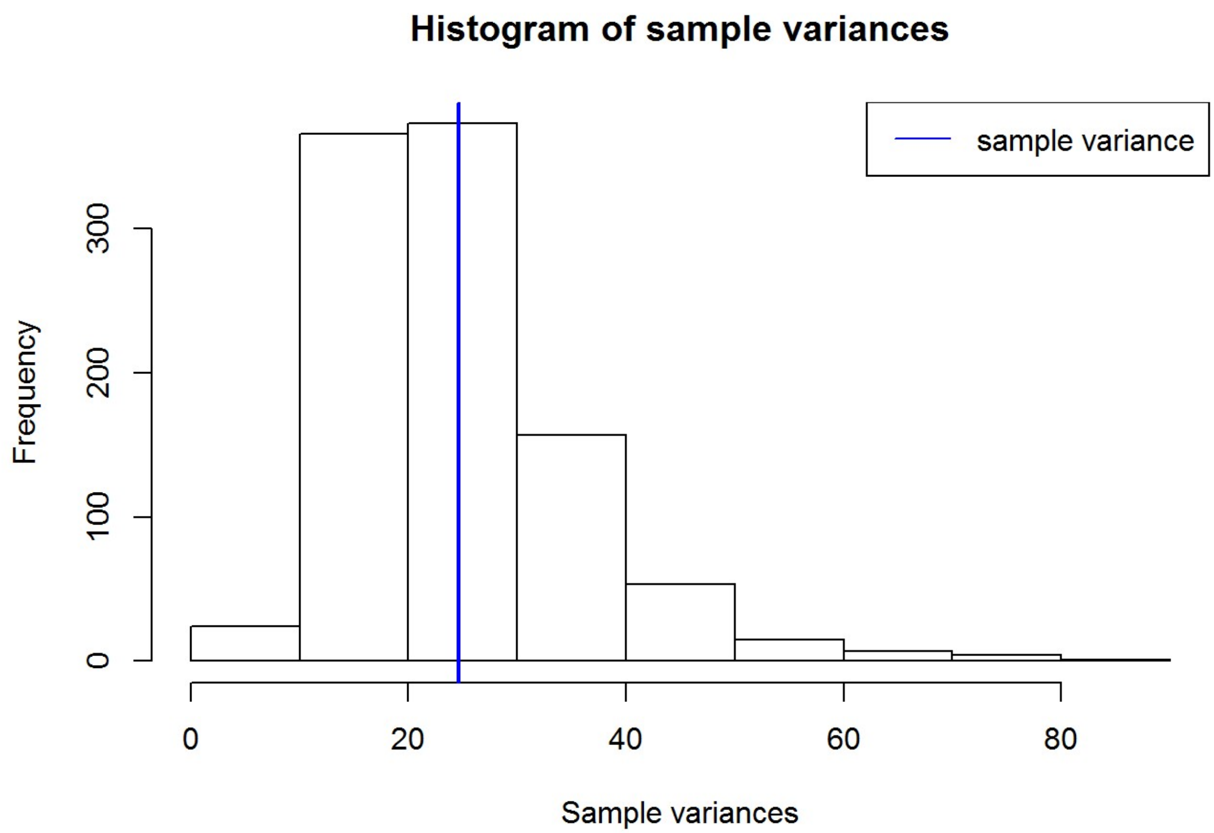
```
hist(means, xlab = 'Sample means', main='Histogram of density of sample means')
abline(v = sample_mean, lwd=2, col='blue')
```



Sample Variance vs Theoretical Variance

Now to compare the variances between the simulation and the theoretical. The theoretical variance is known to be $(1/lambda)^2$ in our example $lamba = 0.2$ so the theoretical variance is 25. The sample variance we calculated above by taking the mean of the variance samples collected during the simulation stage. This gave us a distribution of sample variances and we take the expected value of this variance distribution which will approximate the theoretical variance. The sample variance is 24.5958288. Same as the sample mean the sample variance is very close to the theoretical variance. The more data we collect the more concentrated the distribution of sample variances would be about the theoretical variance.

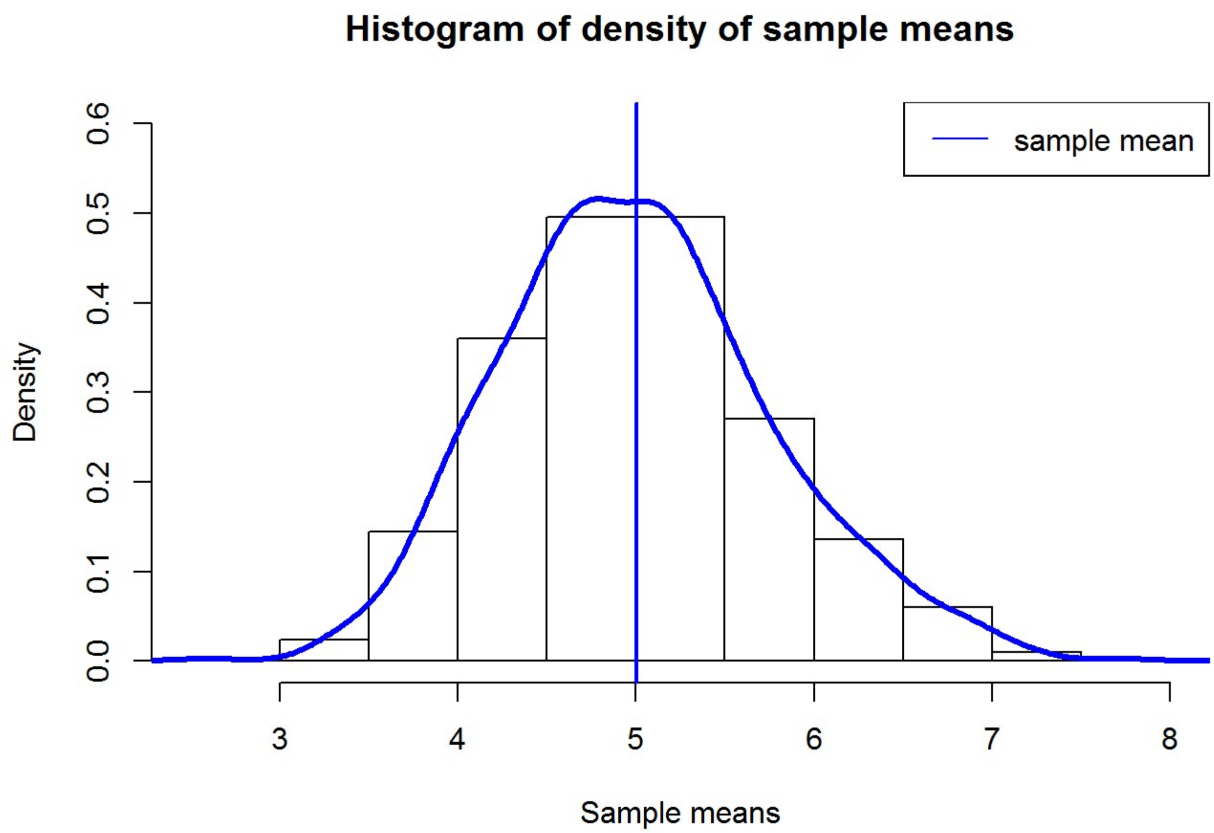
```
hist(variances, xlab='Sample variances', main='Histogram of sample variances')
abline(v = sample_variance, lwd=2, col='blue')
legend('topright', c('sample variance'), lty = 1, col = c('blue'))
```



Distribution

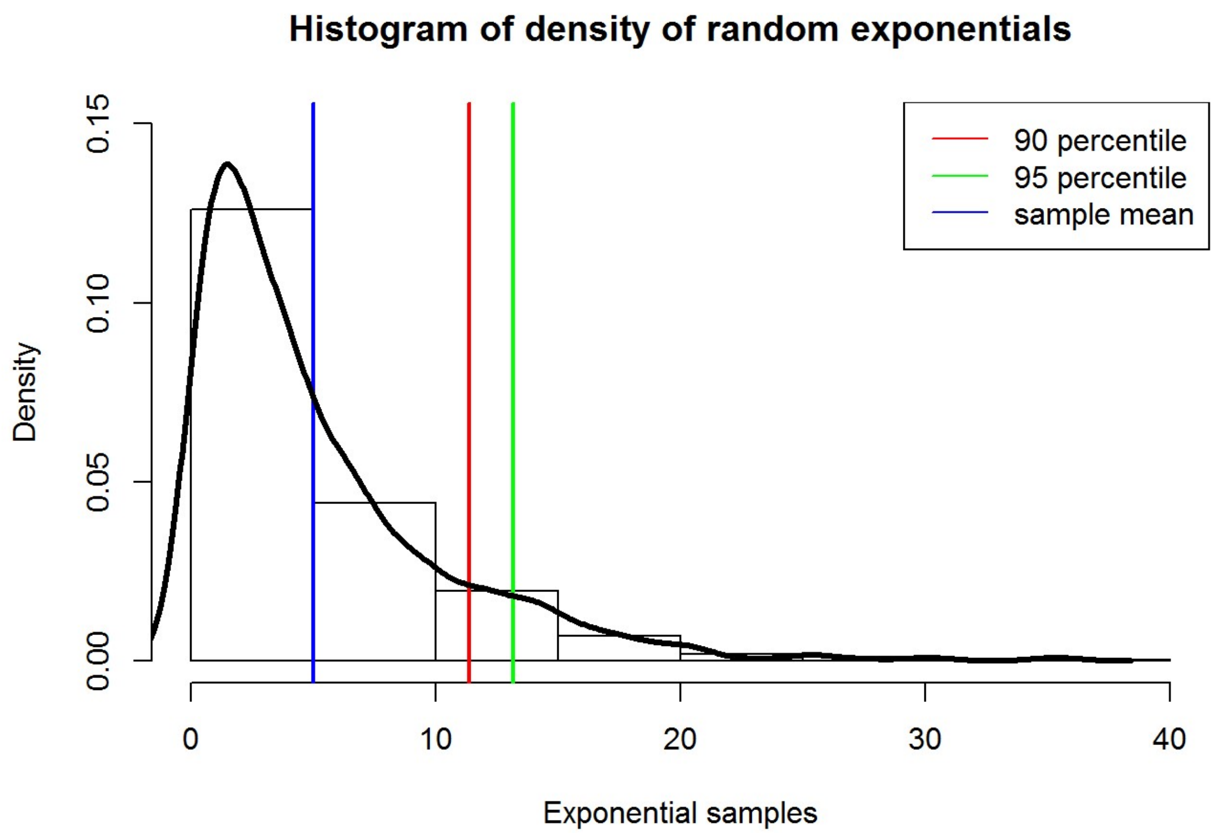
Now we want to see if the exponential distribution is approximately normal. A good way to first see if it is even close is to look at the probability density of our sample means. If the distribution is approximately normal we will see a bell curve looking density. We can show more convincingly that the exponential distribution with a $lambda = 0.2$ is actually a normal distribution. One fact about normal distributions is that if we take the standard normal Z we can convert it into the a non-standard normal by multiplying by sigma and adding the mean ($X = \mu + \sigma Z\ N(\mu, \sigma^2)$). This happens to match perfectly our theoretical mean and variance and we have just shown above that our sample mean and sample variance approximate the theoretical exponential distribution. Therefore the exponential distribution must be a normal distribution because you can go to and from the standard normal distribution.

```
hist(means, probability = T, xlab = 'Sample means', main='Histogram of density
of sample means', ylim = c(0,0.6))
abline(v = sample_mean, lwd=2, col='blue')
lines(density(means), lwd=3, col='blue')
legend('topright', c('sample mean'), lty = 1, col=c('blue'))
```



The above plot looks like a bell curve. It looks approximately centered on the sample mean, 5.0047087 and is symmetric. This is just looking at the distribution of a large collection of averages of 40 exponentials. This is good for trying to find out the population mean and variance. Now let's compare it to the large distribution of random exponential samples. The plot below shows a density histogram of 1,000 exponential samples. The plot also shows that you can convert from the standard normal to this distribution. There are vertical lines drawn showing the 90th, 95th percentiles of the distribution. This was done by taking standard normals 1.28, and 1.645 and converting them to this distribution by adding the mean and multiplying by the standard deviation as described earlier ($X = \mu + \sigma Z \ N(\mu, \sigma^2)$). This shows that the exponential distribution is a normal distribution.

```
exp_samples <- rexp(n, lambda)
hist(exp_samples, probability = T,xlab = 'Exponential samples', main='Histogra
m of density of random exponentials',
     ylim=c(0,0.15))
abline(v = sample_mean, lwd=2, col='blue')
sample_sd = sqrt(sample_variance)
abline(v = sample_mean + 1.28*sample_sd, lwd=2, col='red')
abline(v = sample_mean + 1.645*sample_sd, lwd=2, col='green')
lines(density(exp_samples), lwd=3)
legend('topright', c('90 percentile', '95 percentile', 'sample mean'), lty =
1, col=c('red', 'green','blue'))
```


Histogram of density of random exponentials
90 percentile
95 percentile
sample mean
Density
0.00
0.05
0.10
0.15
0
10
20
30
40
Exponential samples